Medical Expenditure Panel Survey

HOUSEHOLD COMPONENT WEIGHTING AND ESTIMATION ISSUES

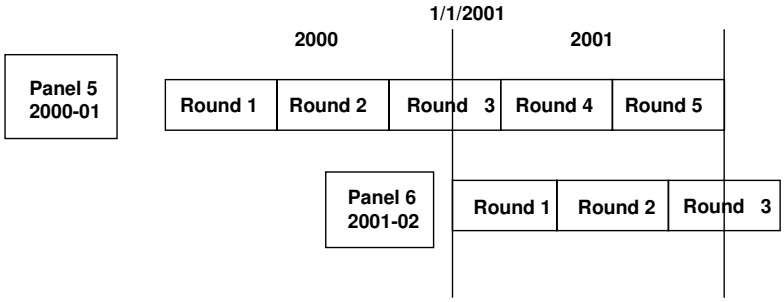
Topics Covered

- Annual person-level estimates
  - Overlapping panels
- Weights and variance estimation variables
  - Weights
  - Variance
- Using annual data from multiple years
  - Time trends
  - Pooling data
- Longitudinal analysis of MEPS panels
  - Two-year period
- Family-level estimation
- Other miscellaneous issues

ANNUAL PERSON-LEVEL ESTIMATES

MEPS-HC Overlapping Panels 2001 Annual File

Panel 5 Year 2 and Panel 6 Year 1



MEPS-HC Annual Files

Year Panel	1997	1998	1999	2000	2001
1 (96-97)	Yr. 2				
2 (97-98)	Yr. 1	Yr. 2			
3 (98-99)		Yr. 1	Yr. 2		
4 (99-00)			Yr. 1	Yr. 2	
5 (00-01)				Yr. 1	Yr. 2
6 (01-02)					Yr. 1

MEPS-HC
Annual Person-Level Estimation

	1997	1998	1999	2000	2001
File Number	HC-020	HC-028	HC-038	HC-050	HC-060
Persons with weight > 0	32,636	22,953	23,565	23,839	32,122
Weighted Persons: All	271.3 million	273.5 million	276.4 million	278.4 million	284.2 million
INSC1231=1*	267.7 million	270.1 million	273.0 million	275.2 million	280.8 million

*Persons in target population at end of year

Observations with person weight=0 have positive family weights. These are persons who are not in the civilian noninstiutionalized population but are family members of sample persons who are in scope for the survey. They are included in the file for family-level estimation.

MEPS-HC Annual Person-Level Estimation (cont'd)

	2002	2003	2004	2005	2006
File Number	HC-070				
Persons with person wt > 0	37,418				
Weighted Persons: All	288.2 million				
INSC1231=1*	284.6 million				

*Persons in target population at end of year

Observations with person weight=0 have positive family weights. These are persons who are not in the civilian noninstiutionalized population but are family members of sample persons who are in scope for the survey. They are included in the file for family-level estimation.

WEIGHTS AND VARIANCE ESTIMATION VARIABLES

MEPS-HC Sample Design

- Each panel is subsample of household respondents for the previous year's National Health Interview Survey (NHIS)
  - NHIS sponsor is National Center for Health Statistics
- NHIS sample based on complex stratified multi-stage probability design
- Civilian noninstitutionalized population.

NHIS Sample Design (1995–2004)

- U.S. partitioned into 1,995 Primary Sampling Units (PSUs) (counties or groups of adjacent counties)
- PSUs grouped into 237 design strata
  - 358 PSUs sampled across strata
- Second Stage Units (SSUs)
  - Clusters of housing units
  - Oversample of SSUs with large Black/Hispanic populations
- MEPS based on subsample of about 200 PSUs from NHIS

Oversampling in MEPS-HC

- Every year: Blacks and Hispanics
  - Carryover from NHIS
- 1997: Selected subpopulations
  - Functionally impaired adults
  - Children with activity limitations
  - Adults 18-64 predicted to have high medical expenditures
  - Low income
  - Adults with other impairments
- 2002 and beyond:
  - Asians
  - Low income
  - Additional oversampling of blacks in 2004

Estimation from Complex Surveys

- Estimates need to be weighted to reflect sample design and survey nonresponse.
  - Unweighted estimates are biased.
- Use appropriate method to compute standard errors to account for complex design.
  - Assuming simple random sampling usually underestimates sampling error.

Development of Person Weights

- **Base Weight (NHIS)**
  - **Compensates for oversampling and nonresponse**
- **Adjustments for**
  - **Household nonresponse (MEPS Round 1)**
  - **Attrition of persons (subsequent Rounds)**
  - **Poststratification (Census population estimates)**
  - **Trimming of extreme weights**
- **Final Person Weight**
  - **Small proportion of cases assigned weight = 0**

- Development of person weights is a multi-stage process.
- Starts with NHIS weight and adjusts to compensate for households and persons without responses to MEPS for the entire year.
- Characteristics used to adjust weights were related to both the likelihood of nonresponse and survey measures.
- Weights adjusted to match Census population estimates on selected characteristics, including sex, age, race/ethnicity, geographic area.
- Variable name for person-level weight is WTDPER96.
- Several person-level weights evolved as more data became available (see variable matrix). Weights on earlier files are obsolete, the most current person-level weight is WTDPER96, which accounts for deaths and nursing home admissions during the year.

Distribution of MEPS-HC Sample Person Final Weights

	1997	1998	1999	2000	2001
Average	8,312	11,917	11,730	11,679	8,849
Minimum	299	321	307	454	336
Maximum	68,518	84,587	80,062	78,157	67,537
Variable Name	WTDPER97	WTDPER98	PERWT99F	PERWT00F	PERWT01F

Distribution of MEPS HC Sample Person Final Weights (cont'd)

	2002	2003	2004	2005	2006
Average	7,702				
Minimum	367				
Maximum	46,766				
Variable Name	PERWT02F				

Types of Basic Point Estimates

- Means
- Proportions
- Totals
- Differences between subgroups

Variance Estimation

- **Basic software procedures assume simple random sampling (SRS)**
  - **MEPS not SRS**
  - **Point estimates correct (if weighted)**
  - **Standard errors usually too small**
- **Software to account for complex design using Taylor Series approach**
  - **SUDAAN (stand-alone or callable within SAS)**
  - **STATA (svy commands)**
  - **SAS 8.2 (survey procedures)**
  - **SPSS (new complex survey features in 13.0)**

- MEPS is not a simple random sample, and standard errors for many estimates from the survey will be larger than expected from a simple random sample.
- SUDAAN and STATA are two commonly used software packages to estimate standard errors from surveys with complex sample designs like MEPS.
- Latest version of SAS (V8.2) will account complex design only if macro SMSUB is used. This macro can be downloaded from the SAS Web site.
- Other analysis software is discussed on the following Web site http://www.fas.harvard.edu/~stats/survey-soft/survey-soft.html and includes

Bascula from Statistics Netherlands.

CENVAR from U.S. Bureau of the Census.

CLUSTERS from University of Essex.

Epi Info from Centers for Disease Control.

Generalized Estimation System (GES) from Statistics Canada.

IVEware (beta version) from University of Michigan.

PCCARP from Iowa State University.

VPLX from U.S. Bureau of the Census.

Estimation Example: Average Total Expenditures, 2001

- **Weighted mean = $2,555 per capita**
  - **Unweighted mean of $2,400 is biased**
- **SE based on Taylor Series = 55**
  - **SAS V8.2: PROC SURVEYMEANS**
  - **SUDAAN: PROC DESCRIPT**
  - **Stata: svymean**
- **SE assuming SRS = 41 (too low)**
  - **SAS V8.2: PROC UNIVARIATE or PROC MEANS**

- This example was programmed using SAS V8.2; it did not use the SAS macro SMSUB.
- This slide shows that the standard error on total expenditures for 1998 are equivalent when the whole sample is used.

Example (Point estimates and SEs): SAS V8.2

- **proc surveymeans data=work.h60 mean;**
  stratum varstr01;
  cluster varpsu01;
  weight perwt01f;
  var totexp01;

Example (Point estimates and SEs): SUDAAN (SAS-callable)

- First need to sort file by varstr01 & varpsu01
- proc descript data=work.h60 filetype=SAS design=wr;
  nest varstr01 varpsu01;
  weight perwt01f;
  var totexp01;

Example (Point estimates and SEs): Strata

```
Svyset [pweight=perwt01f],
   strata(varstr01) psu (varpsu01)
   svymean(totexp01)
```

Analysis of Subpopulations

- Analyzing files that contain only a subset of MEPS sample may produce error messages or incorrect standard errors.
- Each software package has capability to produce subpopulation estimates from entire person-level file.
- See Fact Sheet:
  - "Computing Standard Errors for MEPS Estimates"

Assessing Precision/Reliability of Estimates

- Sample Sizes
- Standard Errors/Confidence Intervals
- Relative Standard Errors (RSE)
  - standard error of estimate ÷ estimate

Example: Average total expenses per capita, 2001

- Sample Size = 32,122
- Estimate = $2,555
- Standard Error = 55
- 95% Confidence Interval: (2447, 2663)
- Relative Standard Error (RSE) or Coefficient of Variation (CV) = 55 ÷ 2555 = .022 = 2.2%

Types of Basic Point Estimates: Examples

- Means
  - Annual per capita expenses in 2001 = $2,555
- Proportions
  - Percent with some health expenses in 2001 = 85.4%
  - Two methods to generate estimates:
    - percents obtained from frequency tables
    - means of dichotomous variable
- Totals
  - Total expenses in 2001 = $726.4 billion
  - Total number of persons (sum of weights)
- Differences between subgroups

USING MEPS-HC TO EXAMINE TRENDS OVER TIME

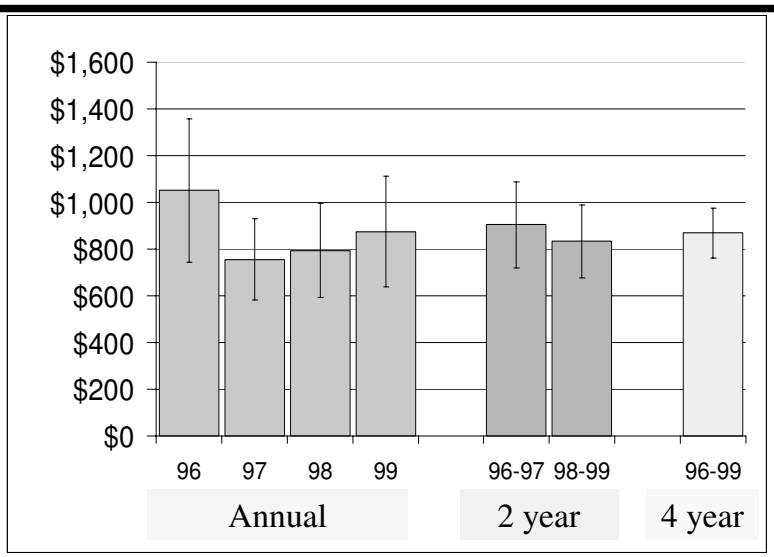
Interpreting Trends Using MEPS-HC Annual Data

- **Statistical significance**
  - **Differences across years may reflect sampling error.**
  - **Expense estimates sensitive due to skewed distribution.**
- **Length of time analyzed**
  - **Interpret large shifts in short periods with caution.**
  - **Consider timing of policy changes.**
  - **Changes to MEPS methodology in documentation.**
  - **Consider standardizing medical expenses by CPI when analyzing trends.**

Alternatives to Analyzing Trends Based on Annual Data

- Pooling years
  - E.g., 1996–97 versus 1998–99
- Examine moving averages
  - E.g., 1996–97, 1997–98, 1998–99, etc.
- Modeling techniques
  - Use several consecutive years
  - Test fit of specified patterns over time

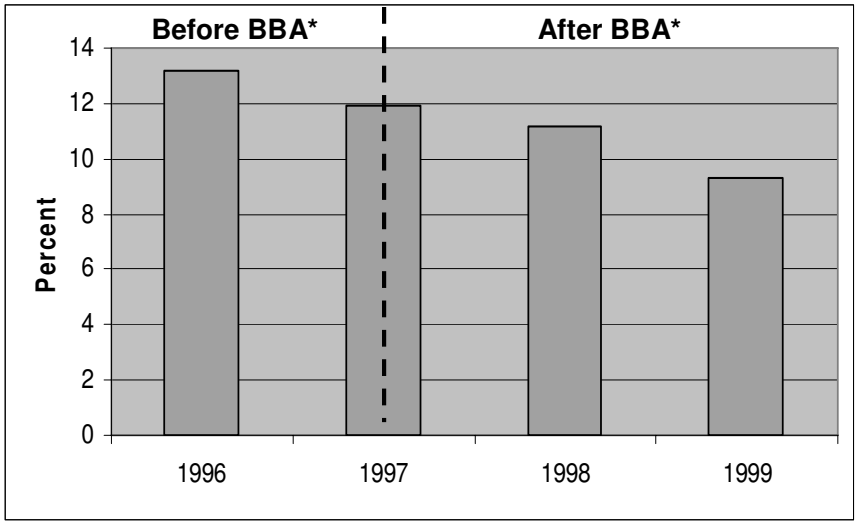
Estimated Mean Expenditures (95% CI): Children Ages 0–5



Note: Pooled 2 year and 4 year estimates in 1999 dollars.

A lot more variability in estimates for children. Note the drop from 96 to 97 (p-value is about .09). Pooled estimates stabilize the estimates.

Percent of Persons 65 and Over with Home-Care Expenditures



*Balanced Budget Amendment

POOLING MULTIPLE YEARS OF MEPS-HC DATA

Reasons for Pooling

- Reduce standard error of estimate(s)
- Stabilize trend analyzes
- Enhance ability to analyze small subgroups

Minimum Sample Sizes

- CFACT Standards
  - Minimum unweighted sample of 100
  - Flag estimates with RSE > 30%
- Confidence intervals become problematic with small samples and/or highly skewed data
  - Consider larger minimum sample sizes for highly skewed variables
  - Analysts may be comfortable with smaller minimums for less skewed variables
- ASA papers (handouts)
  - Machlin, Zodet, and Nixon (Pooling)
  - Yu and Machlin (Skewness)

Example: Annual Sample Sizes (Unpooled)

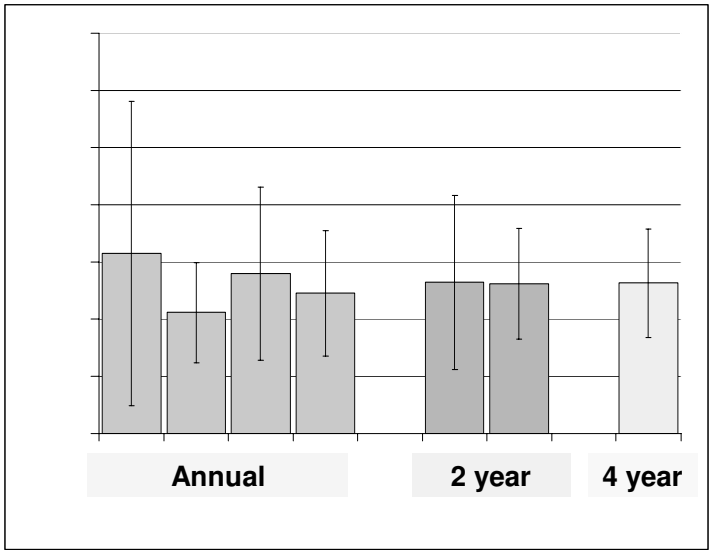
Year	Total Population	Children 0-5	Asian/PI Children* 0-5
1996	21,571	2,018	58
1997	32,636	3,082	78
1998	22,953	2,114	82
1999	23,565	2,156	93

*** Sample sizes do not meet AHRQ minimum requirement (n=100) to publish estimates.**

Pooled Sample Sizes

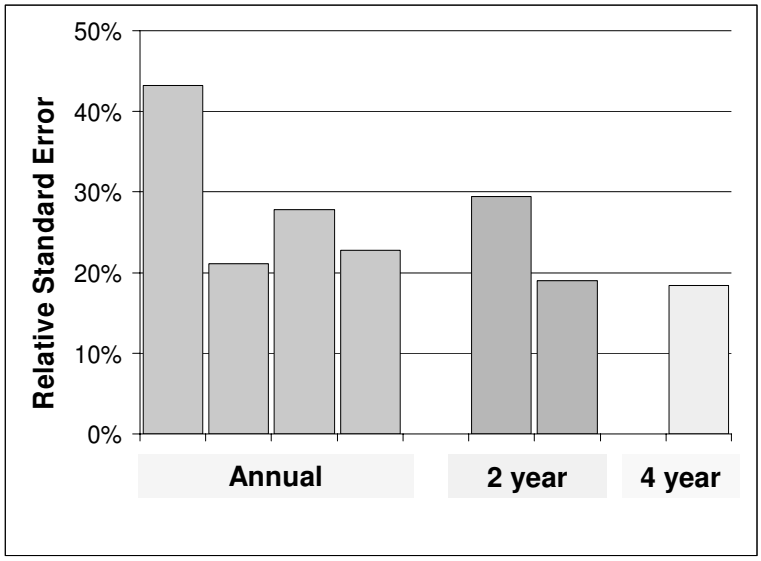
Years	Total Population	Children 0–5	Asian/PI Children 0–5
1996-1997	54,207	5,100	136
1998-1999	46,518	4,270	175
1996-1999	100,725	9,370	311

Estimated Mean Expenditures (95% CI): Asian/PI Children 0–5



Means fluctuate and CIs are very wide for API children. Y-axis suppressed.

Relative Standard Errors for Estimated Mean Expenditures: Asian/PI Children 0–5



RSEs are extremely high for annual estimates. Pooled estimates are still high -- about 30% for 96 to 97 and 20% for 98 to 99. Pooled four-year slightly under 20%. Publishable by AHRQ standards but not extremely precise.

Creating a Pooled File for Analysis (1996–2002)

- Need to work with Pooled Estimation File (HC-036) when 1+ years being pooled are in the group of years from 1996 and 2001
  - Stratum and PSU variables obtained from HC-036 for 1996-2002
  - Stratum and PSU variables properly standardized for pooling years from 2002 onward
- Documentation provides instructions on how to properly create pooled analysis file

Creating Pooled Files: Summary of Important Steps

- Rename analytic and weight variables from different years to common names.
  - Expenditures: TOTEXP99 & TOTEXP00 = TOTEXP
  - Weights: PERWT99F & PERWT00F = POOLWT
- Divide weight variable by number of years pooled to produce estimates for "an average year" during the period.
  - Keep original weight value if estimating total for period.
- Concatenate annual files.
- Merge variance estimation variables from HC-036 onto file.
  - Strata variable: STRA9602
  - PSU variable: PSU9602

Estimates from Pooled Files

- Produce estimates in analogous fashion as for individual years
- Estimates interpreted as "average annual" for pooled period
- Example: Pooled 1996-99 data
  - The average annual total health care expenditures for Asian/Pacific Islander children under 6 years of age during the period from 1996–1999 was $525 (SE=97).

Pooling Annual Data: Lack of Independence across Years

- Legitimate to pool data for persons in consecutive years.
  - Each year constitutes nationally representative sample.
  - Pooling produces average annual estimates.
  - Stratum and PSU variables sufficient to account for lack of independence between years.
- Lack of independence actually begins with first stage of sample selection.
  - Same PSUs are used to select each MEPS panel.
- See HC-036 documentation: Section 4.0.

Pooling Annual Data: Average Annual Expenditures

- Consider standardizing expenses by Overall Consumer Price Index (CPI) when analyzing multiple years.
- Example: When pooling 1999 & 2000, adjust 1999 to 2000 dollars by multiplying TOTEXP99 * (172.2 / 166.6).

Year	Overall CPI
1996	156.9
1997	160.5
1998	163.0
1999	166.6
2000	172.2
2001	177.1
2002	179.9

Longitudinal Analysis of MEPS-HC Panels

MEPS-HC Longitudinal Analysis: Panel 4: 1999-2000

MEPS Longitudinal Analysis

- National estimates of person-level changes over two-year period
  - two-year period is relatively short
- Examine characteristics associated with changes
  - mainly round 1 data

Variables That May Change between Years or Rounds

- Insurance coverage
  - Monthly indicators (24 measures)
  - Annual summary (2 measures per person)
- Health status
  - Each round (5 measures)
- Having a usual source of care
  - Rounds 2 & 4 (2 measures)
- Use and expenditures
  - Annual (2 measures per person)

MEPS Longitudinal Files*

MEPS Panel	Yrs Covered	PUF Number
1	1996-97	HC-023
2	1997-98	HC-035
3	1998-99	HC-048
4	1999-00	HC-058
5	2000-01	HC-065

* Available as of December 2004

Creating Longitudinal Files (Panel 4) : Summary of Important Steps

- Select Panel 4 records from annual files
  - 1999 (PUF HC-038)
  - 2000 (PUF HC-050)
- Obtain MEPS Longitudinal File (HC-058)
  - Contains weight and variance estimation variables
  - Contains variable indicating whether complete data are available for 1 or both years of panel
- Link using DUPERSID

Longitudinal Weight

- **Variable Name: LONGWTP#**
- **Produces estimates for persons in civilian noninstitutionalized population in two consecutive years when applied to persons participating in both years of a given panel (YRINDP# = 1)**

Examples: Insurance Coverage

- Of those without insurance at any time in 1999,
  - Estimated 76.9% (SE=1.6) also uninsured throughout 2000
- Of those with insurance in 1999,
  - Estimated 3.5% (SE=0.3) were uninsured throughout 2000
- Estimated 8.2% (SE=0.4) of the population had no insurance throughout 1999-2000

Examples: Health Care Expenditures

- Of those with some expenses in 1999,
  - Estimated 88.9% (SE=0.3) also had expenses in 2000
- Of those with no expenses in 1999,
  - Estimated 52.4% (SE=1.3) had some expenses in 2000
- Of top 5% of spenders in 1996, 30% retain this position in 1997.
  - Based on longitudinal analysis of Panel 1

Family-Level Estimation

Family-Level Estimation

- Need to roll up persons to families
  - MEPS vs. CPS definitions
  - Any time during year or December 31
  - Instructions in person file documentation
- Average number of persons per family = 2.4
- Use appropriate family weight variable

MEPS Annual Files: MEPS Annualized Family Sample Sizes

	1997	1998	1999	2000	2001	2002
File Number	HC-020	HC-028	HC-038	HC-050	HC-060	HC-070
Families (unwtd)	13,087	9,023	9,345	9,515	12,852	14,828
Weighted	112.2 million	113.4 million	114.6 million	116.3 million	118.8 million	121.0 million
Family Weight Variable Name	WTFAMF97	WTFAMF98	FAMWT99F	FAMWT00F	FAMWT01F	FAMWT02F

Family-Level Example

- 2001 average total expenses per family
- Estimates based on families in scope at any time during year

Family size	Estimate	SE
All	$6,029	131
1	$4,191	215
2	$7,405	277
3	$6,616	268
4	$6,075	278
5+	$7,518	389

Other Miscellaneous Estimation Issues

Medical Event as Unit of Analysis

- Can use event files to estimate average expense per event
- Examples: In 2001,
  - mean facility expense per inpatient stay was $6,629 (SE=263).
  - mean expense per office visit to a medical provider was $114 (SE=2)

Special Supplements

- Self Administered Questionnaire (SAQ)
  - Use SAQ weight
- Parent Administered Questionnaire (PAQ)
  - 2000 only
  - Use PAQ weight
- Diabetes Care Survey (DCS)
  - Use DCS weight
- Variables on person-level files
  - Consult documentation for appropriate weight